Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 15 (2011) 1742 – 1747

Procedia Engineering

www.elsevier.com/locate/procedia

Advanced in Control Engineering and Information Science

Multi-Dimension Sandpile Space: a new theory of representation and reasoning for experience knowledge

Yu JIN*, Wen-jun HOU, Chang-hua LIU

Automation School, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract

Fuzzy experience knowledge cannot be explained quantitatively with mathematical formula or certain rules. In order to solve this problem, a new theory of Multi- Dimension Sandpile Space (MDSS) was proposed in this paper with analogy to the Sandpile Model in self-organization theory. MDSS theory has the ability of dealing with uncertain or fuzzy knowledge by combining the virtues of fuzzy logic theory and neural networks, the characteristic of continual learning of case-based reasoning (CBR), and avoided the unceasing inflation of the size of information in database when using CBR. Finally, the simulation results of an example are given to show the change of the size of database by using MDSS theory.

Selection and/or peer-review under responsibility of [CEIS 2011]

Keywords: experience knowledge; knowledge representation; knowledge reasoning; Multi-Dimension Sandpile Space; user-context-items

1. Introduction

Knowledge is the sum of understanding and experience about natural and society in the practice of reforming the objective world, and also the sorting and processing of information. From Epistemology and Cognitive Psychology perspective, Polanyi [1] divided knowledge into Explicit Knowledge and Tacit Knowledge in 1966, providing a foundation for understanding knowledge and knowledge activities. Tacit

* Corresponding author. Tel.: +86-010-62281505.
E-mail address: goldfisher.jin@gmail.com.

doi:10.1016/j.proeng.2011.08.325

Knowledge is knowledge that cannot be easily understood and grasped. It is long-term accumulated by human, reflecting one's experience. The representation and reasoning for experience knowledge is an important part in knowledge management.

There are already some researches about fuzzy knowledge representation and reasoning at present, such as CBR (case-based reasoning) [2], rough set [3], neural networks [4], Slope One [5], and so on. CBR (case-based reasoning) [2] is a reasoning method based on historical records of experience knowledge. However, the maintenance of the database is difficult. Applications of rough set theory and neural networks algorithm have solved some problems of uncertainty, but they cannot ensure the real-time updating. Slope One algorithm is suitable for the situation with user-item model. It can do nothing when user is going to make a choice of items under a particular context.

Many previous knowledge recommendation algorithms were applied to the situation which can be described well with the user-items model, in which user and items are directly related and items are discrete, as shown in Fig. 1(a). However, there is some kind of experience knowledge which belongs to user-context-items model, in which user will choose one item under certain context. Although items are discrete too, context is continuous. And under the same context, several items may have possibility to be chosen by user, as shown in Fig. 1(b).

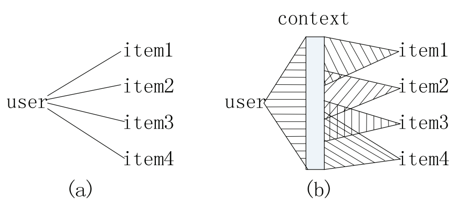

Fig. 1. (a) user-items model; (b) user-context-items model

This paper put forward a new theory to overcome the disadvantages of the above methods. It has the ability of dealing with uncertain or fuzzy knowledge by combining the virtues of fuzzy logic theory and neural networks, the characteristic of continual learning of case-based reasoning (CBR), and avoided the unceasing inflation of the size of information in database when using CBR.

Nomenclature

maxH	maximum height of a sandpile
$item_i$	the i^{th} item which is used to represent specific experience knowledge
SP_i	the i^{th} sandpile which is used to represent the i^{th} item in sandpile space
c_i	represents some specific context
$AddSand(c_i)$	represents the behaviour that making a choice of item under the context of c_i, also means putting one pinch of sand on sandpile at the point c_i

2. Multi-Dimension Sandpile Space Theory

Sandpile Model was used to give a concrete description to the formation and features of self-organization criticality by Bak et al [6]. Self-organization criticality theory focused on the state of avalanche in Sandpile Model. Nevertheless, the purpose of this paper is to research the process of the sandpile formation in Sandpile Model. And according to this, Multi-Dimension Sandpile Space (MDSS) was put forward and used in Tacit Knowledge representation and reasoning.

2.1. The forming process of sandpile

One item may be chosen by user under a range of context, and there are many kinds of context. We consider one kind of context as a dimension in multi-dimension space, and use sandpile in this multi-dimension space to represent user's experience knowledge. The height of sandpile is used to measure the possibility of one item chosen by user.

According to this, we set the rules below on the forming process of sandpile.

Rule 1: Before $AddSand(c_0)$, if the height at the point c_0 is represented as h_0 and $h_0 \geq maxH$, then $AddSand(c_0)$ have no influence on this sandpile.

Rule 2: Before $AddSand(c_0)$, if the height at the point c_0 is represented as h_0 and $h_0 < maxH$, then there exists a function which can change the height at the point c_0 from h_0 to h_1 after $AddSand(c_0)$. The equation is as follows.

$$h_1=h_0+f_h(h_0) \tag{1}$$

Rule 3: If the maximum height of the sandpile is at the point c_h and the value is h, then the height of the sandpile which is represented as h_a at the point c_a can be calculated by the following equation.

$$h_a=f_a(c_h, h, c_a) \tag{2}$$

If context contains only one dimension, then a typical forming of sandpile can be shown in Fig. 2(a)~(d).

Fig. 2 can be explained as follows: First, perform $AddSand(c_1)$ (as shown in Fig. 2(a)) and the initial height of the sandpile is h_1; Then perform $AddSand(c_1)$ at the same point (as shown in Fig. 2(b)) and after that, the height at the point c_1 has been changed to $h_2=h_1+f_h(h_1)$; Then perform $AddSand(c_2)$ at the point c_2 (as shown in Fig. 2(c)) and the height at the point c_2 has been changed from $h_3=f_a(c_1, h_2, c_2)$ to $h_4=h_3+f_h(h_3)$; Finally, perform $AddSand(c_3)$ at the point c_3 (as shown in Fig. 2(d)) and the final shape of the sandpile is shown in Fig. 2(d) because of $h_6=h_5+f_h(h_5)$ and $h_6>maxH$.

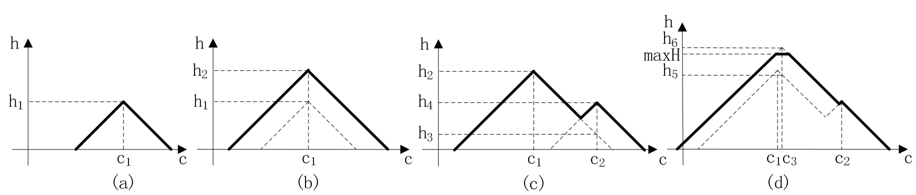

Fig. 2. A typical forming process of a sandpile

2.2. The choice of items

Experience knowledge can be represented by using Sandpile Space. There may be several local maximum values in one sandpile, and different items should be expressed by corresponding sandpile. If the sandpile in Fig. 2 represents $item_i$, then Fig. 2 can be understood as choosing $item_i$ twice at the point c_1 and choosing $item_i$ once at the point c_2 and c_3 separately. The final shape means $item_i$ has the possibility to be chosen by users under the whole sandpile.

Because the past experience knowledge may not apply to the current situation, there should be a self-correcting mechanism in the knowledge base system to update the experience knowledge.

Rule 4: While performing $AddSand(c_i)$ on SP_1, if the point c_i is also covered under SP_2, then SP_2 need to be updated too. The height of SP_1 at the point c_i is represented as h_{10}, and the height of SP_2 at the same point is represented as h_{20}. Then after performing $AddSand(c_i)$, the height of SP_1 and the height of SP_2 have been changed to $h_{11}= h_{10}+f_h(h_{10})$ and h_{21} separately. The equation of h_{21} is as follows.

$$h_{21}= h_{20}-f_c(h_{20}, h_{11}-h_{10}) \tag{3}$$

Assuming context contains two dimensions c_1 and c_2, we made two sandpiles SP_1 and SP_2 as shown in Fig. 3 to visually explain the updating of sandpile. The number in Fig. 3 means the height of the sandpile at the corresponding position. Assuming $c=(3,3)$ and after performing $AddSand(c)$ on SP_2, the results are shown in Fig. 3(b) and Fig. 3(d).

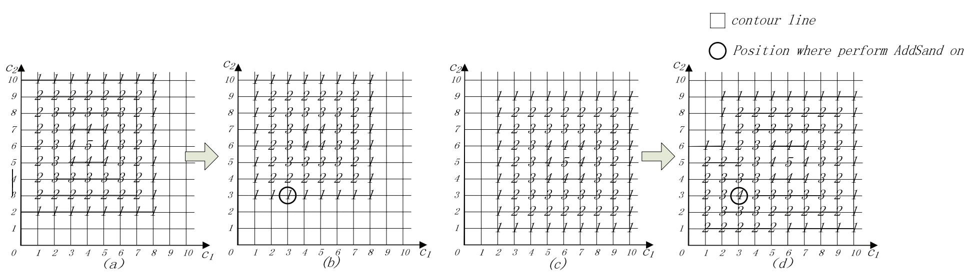

Fig. 3. A diagram of self-correcting mechanism and the overlap of sandpiles

Experience knowledge has uncertainty, so overlaps often appear among sandpiles. As shown in Fig. 3(a) and Fig. 3(c), the height of SP_1 is $h_1=4\neq0$ and the height of SP_2 is $h_2=3\neq0$ at the point $c=(4,5)$. So it is necessary to rank items according to their possibility to be chosen under some specific context. In this paper, we consider the height of sandpile as the criteria for ranking. As shown in Fig. 3(a) and Fig. 3(c), $h_1>h_2$ at the point $c=(4,5)$, so $item_1$ has more possibility to be chosen by users than $item_2$ at the point c.

2.3. Multi-Dimension Sandpile Space

The above description is a simple 2-dimension sandpile space. According to this, we extended this to Multi-Dimension Sandpile Space. The definition of MDSS is as follows.

N-dimension sandpile space consists of n c_i-axises (i=1~n) and one h-axis, where the value on c_i-axis means context and the value on h-axis means the possibility of item to be chosen. With the increasing of the value on h-axis, the possibility of item to be chosen becomes greater and greater. Each item *corresponds to one sandpile. Each time an item is chosen, the corresponding sandpile needs to be updated (as shown in Fig. 2 and Fig. 3).*

According to the foregoing description of MDSS, we know that there are two processes when adding new experience knowledge. They are the process of raising its own height of sandpile and the process of updating the height of other nearby sandpiles (see section 2.1 and 2.2). When user needs to judge which item should be chosen under some specific context according to the existing experience knowledge, the height of all sandpiles under this context needs to be calculated and compared with each other. The higher the height of sandpile is, the higher the possibility of the item being chosen is (see section 2.2).

3. Experimental Results

When representing experience knowledge by using MDSS, it only needs to store each local maximum height of each sandpile and the coordinates of corresponding context.The following example is used to explain the effect of storage by using MDSS.

Suppose that context contains nine dimensions, represented as $c_1 \sim c_9$. Under this context, users have six items to choose, represented as $item_1 \sim item_6$. Which item should be chosen depends on the value of $c_1 \sim c_9$. The range of the value of each dimension in context for each item is shown in Table 1.

Table 1. The range of the value of each dimension in context for each item

items	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
$item_1$	0.0~0.4	0.4~0.7	0.4~0.6	0.6~1.0	0.4~0.8	0.7~1.0	0.6~0.9	0.5~0.8	0.5~0.6
$item_2$	0.2~0.6	0.0~0.4	0.6~1.0	0.0~0.8	0.6~1.0	0.8~0.9	0.5~0.8	0.6~0.9	0.4~0.8
$item_3$	0.3~0.7	0.5~0.9	0.4~0.9	0.4~0.7	0.4~0.8	0.7~0.8	0.7~1.0	0.2~1.0	0.0~0.9
$item_4$	0.3~0.7	0.6~1.0	0.1~0.4	0.2~0.6	0.0~0.4	0.1~0.6	0.4~0.7	0.1~0.5	0.2~0.6
$item_5$	0.6~1.0	0.1~0.5	0.0~0.3	0.5~0.9	0.2~0.6	0.1~0.6	0.2~0.5	0.4~0.7	0.6~0.9
$item_6$	0.6~0.9	0.2~0.6	0.4~0.7	0.4~0.9	0.6~0.8	0.0~0.4	0.0~0.6	0.0~0.1	0.5~1.0

For simplifying the calculation, we set the initial values as follows.

$$f_h(h)=\begin{cases}0.4, & h=0\\ h/2, & h>0\end{cases}, f_a(c_h, h, c_a)=h-(c_h-c_a)*tan(PI/6),\ maxH=1.0,\ f_c(h, h_1-h_0)=(h+h_1-h_0)/n.$$

The value of each dimension in context is formed out of uniform distribution (correct to 0.001), and each item has an equal chance to be chosen. After inputting 100,000 experience knowledge continuously, the changes of the size of knowledge base can be expressed as shown in Fig. 4.

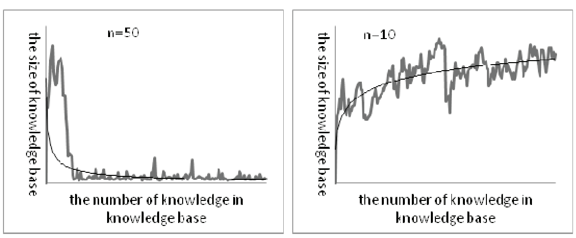

Fig. 4. The changes of the size of knowledge base

When n=50, the value of the function $f_c(h, h_1\text{-}h_0)$ is relatively small, so MDSS has poorer performance of self-correcting mechanism and the sandpile grows too fast. And because the experience knowledge is decentralized in the early stage, there is a great number of sandpiles formed in the same stage. During the process of inputting experience knowledge into knowledge base, the size of knowledge base increased sharply in the early stage; subsequently, it declined gradually until the size is stable among a certain small fluctuations. Another reason of this is that the value of the function $f_c(h, h_1\text{-}h_0)$ is so small that the height of sandpile reached *maxH* too quickly. And while inputting experience knowledge, many decentralized sandpiles began to concentrate to several larger sandpiles.

When n=10, the value of the function $f_c(h, h_1\text{-}h_0)$ is relatively large, so MDSS has stronger performance of self-correcting mechanism and the sandpile grows slow. Therefore, the size of knowledge base increased not as sharply as the case when n=50 in the early stage, and the height of sandpile can hardly reach *maxH*. Similarly, we can also say that larger sandpile can hardly being formed, so there are always many small sandpiles in knowledge base.

4. Conclusion

This paper put forward a new theory called Multi-Dimension Sandpile Space to solve the problem in experience knowledge representation and reasoning. Knowledge base using MDSS can not only deal with uncertain experience knowledge, but also keep knowledge base in an acceptable size when learning new knowledge. And it has an efficient self-correcting mechanism to update knowledge in knowledge base.

However, it is difficult to determine the initial values of MDSS in the early stage. Determining these values needs the designer for this knowledge base to be very familiar with the characteristic of this kind of knowledge, and he also should do many tests. So how to determine these parameters more intelligently should be studied in future.

Acknowledgements

This work is supported by Doctoral Fund of Ministry of Education of China (Grant No. 20090005110015).

References

[1] Polanyi M. The logic of tacit inference. *Philosophy,* Cambridge University Press; 1966; 41: 1-18.

[2] Yang SZ, Ding H, Shi TL. *Knowledge-based diagnostic reasoning*. Tsinghua University Press; 1993.

[3] Pawlak Z. Rough set. *International journal of computer information science* 1982; 11(5): 341-356.

[4] Zhang Liming. *Modeling and application of artificial neural network.* Shanghai: Fudan University Press; 1995. (in Chinese).

[5] DANIEL Lemire, ANNA Maclachlan. Slope One predictors for online rating-based collaborative filtering. *Proceedings of the Fifth SIAM International Conference on Data Mining* 2005; 119: 471-475.

[6]Bak P. *How nature works: the science of self-organized critical*. Springer Press; 1996.